

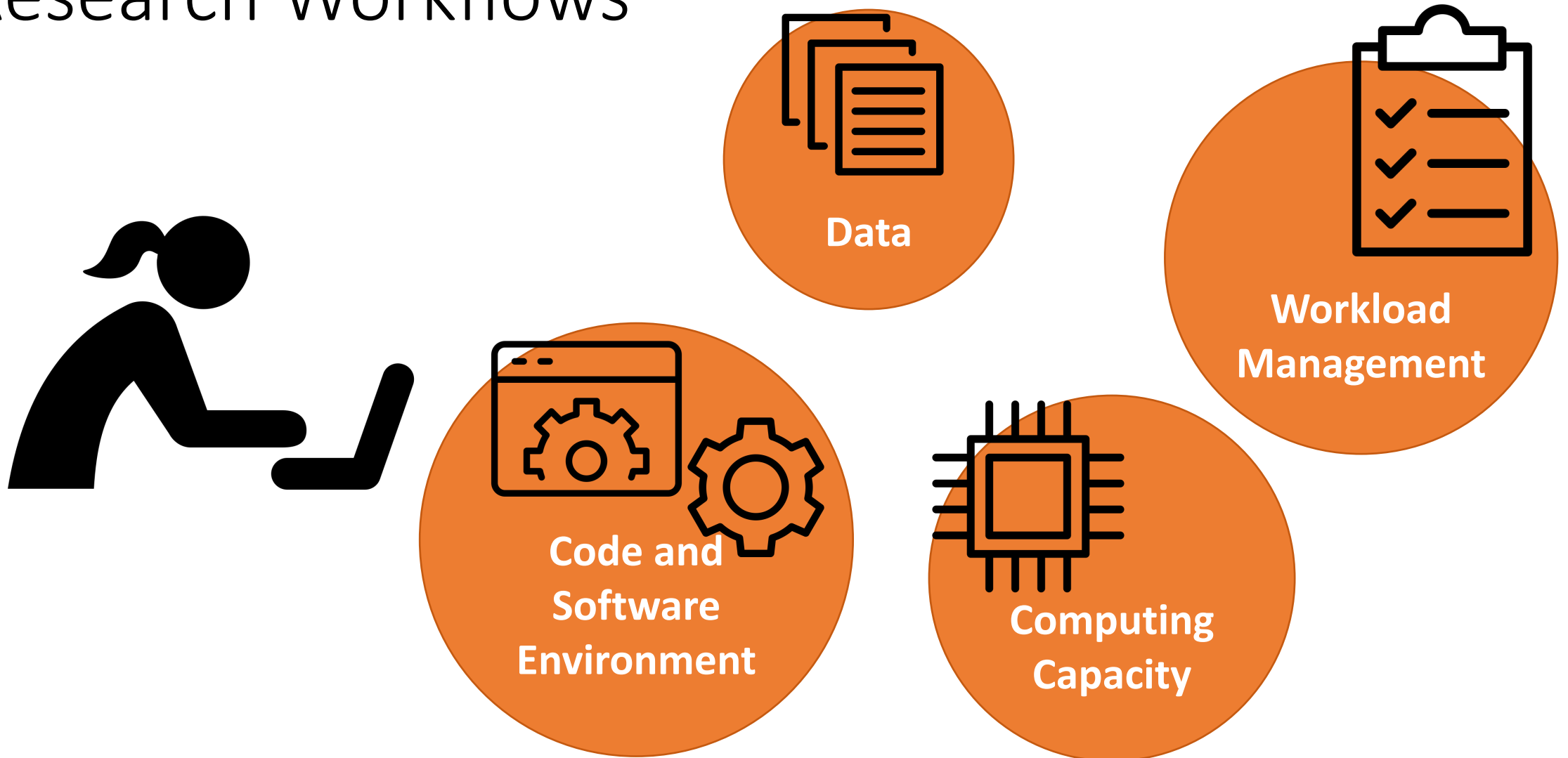
OSG Services for Researchers

Christina Koch

OSG All-Hands Meeting @ Throughput Computing 2023

July 10, 2023

Research Workflows



Workload Management

- How are you going to manage thousands of tasks (jobs)?
- Tools used by OSG Services
 - HTCSS (HTCondor)
 - DAGMan
 - Pegasus
- Important features
 - Managing 1000s of tasks
 - Recording job information

```
executable = compare_states  
arguments = $(state)
```

```
should_transfer_files = YES  
transfer_input_files = us.dat, $(state)  
when_to_transfer_output = ON_EXIT
```

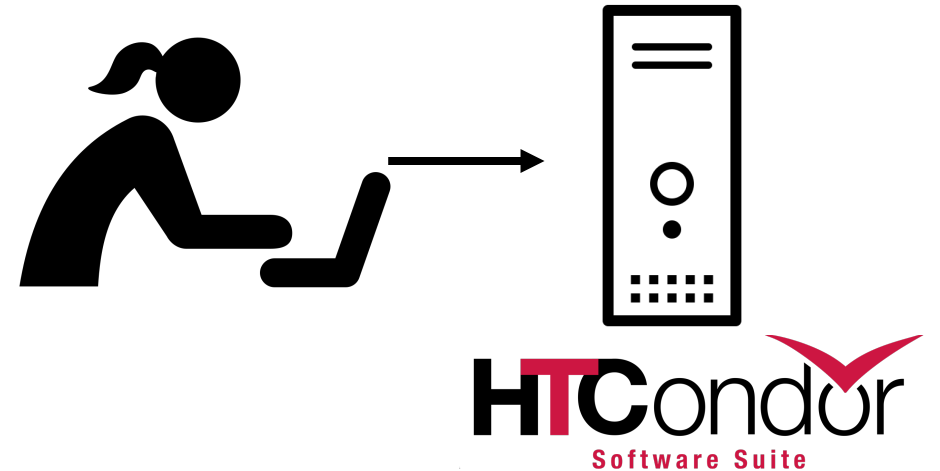
```
log = job.log  
output = job.out  
error = job.err
```

```
request_cpus = 1  
request_memory = 20MB  
request_disk = 20MB
```

```
queue state from datalist.txt
```

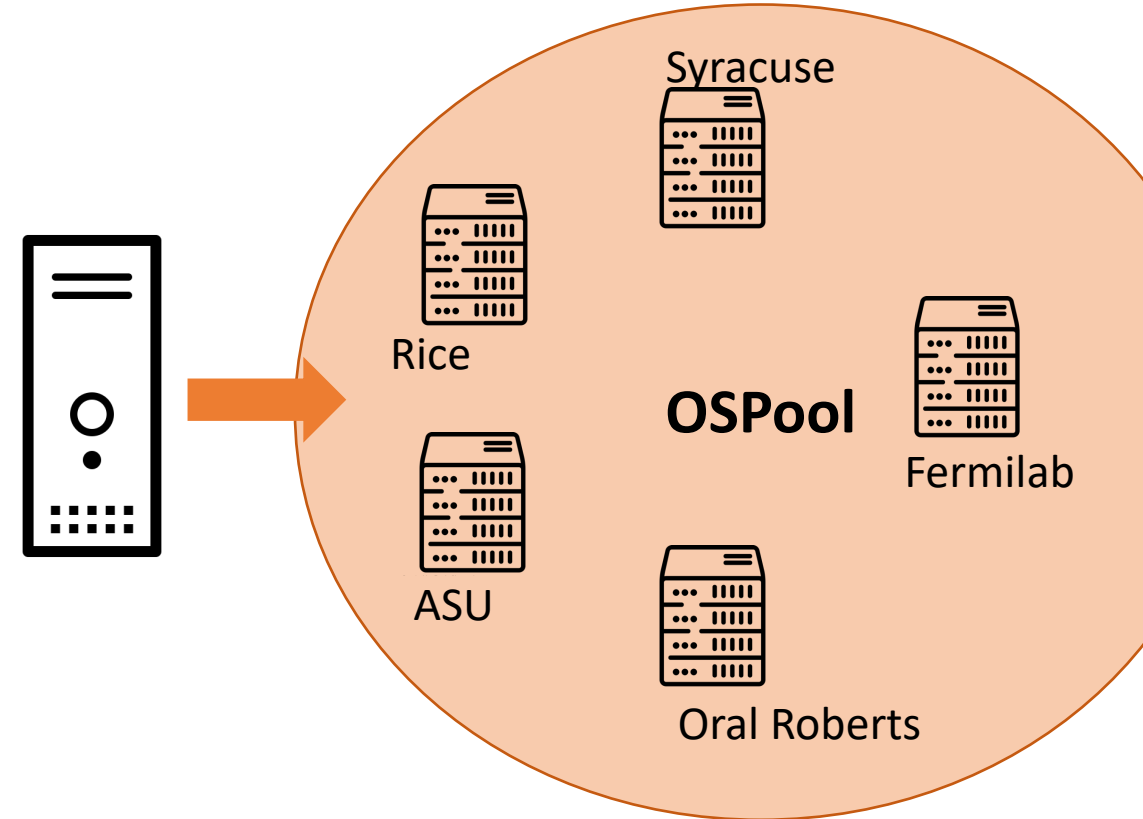
Access Points

- Location (usually a server) where workloads can be placed to run.
 - Has similarities to a “head node” or “login node” on other systems
- Access Points operated by PATH for individual researchers or small-medium collaborations
 - ap2X.uc.osg-htc.org
 - ap4X.uw.osg-htc.org
- Using an AP and related services is free and does not require an allocation. Just request an account!



Computing Capacity: OSPool

- From an Access Point, can access computing capacity.
- Default service for computing capacity is the **OSPool**.
- Composed of contributions from different institutions.
- Expectation is that tasks are isolated to one server (no multi-server MPI-type code)



OSPool: What's Available?

	Total	Owner	Claimed	Unclaimed	Matched	Preempting	Backfill	Drain
x64/AlmaLinux8	716	0	712	4	0	0	0	0
x64/CentOS7	9259	0	9061	198	0	0	0	0
x64/CentOS8	27739	0	27332	406	0	1	0	0
x64/CentOS9	957	0	945	12	0	0	0	0
x64/RedHat7	3793	0	3669	123	0	1	0	0
x64/RedHat8	2058	0	2030	28	0	0	0	0
x64/Rocky8	2182	0	1987	195	0	0	0	0
x64/SL7	3296	0	3249	47	0	0	0	0
x64/Ubuntu18	8	0	0	8	0	0	0	0
x64/Ubuntu20	24	0	1	23	0	0	0	0
Total	50032	0	48986	1044	0			

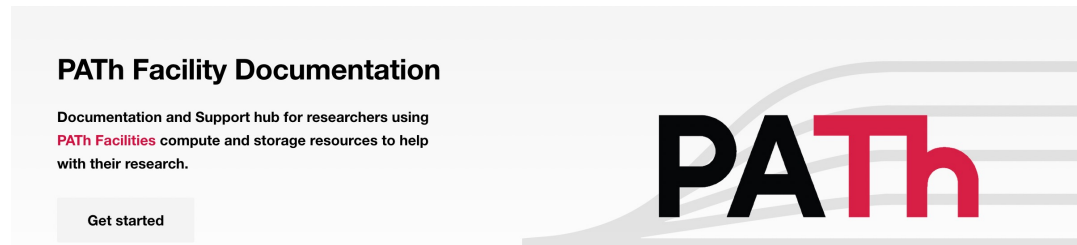
1000s of CPU cores
100s of GPUs

OSPool: What's a Good Fit?

	Ideal Jobs! (up to 10,000 cores across jobs, per user)	Still Very Advantageous!
Cores (GPUs)	1 (1; non-specific type)	<8 (1; specific GPU type)
Walltime	<10 hrs* *or checkpointable	<20 hrs* *or checkpointable
RAM	<few GB	<10s GB
Input	<500 MB	<10 GB
Output	<1 GB	<10 GB

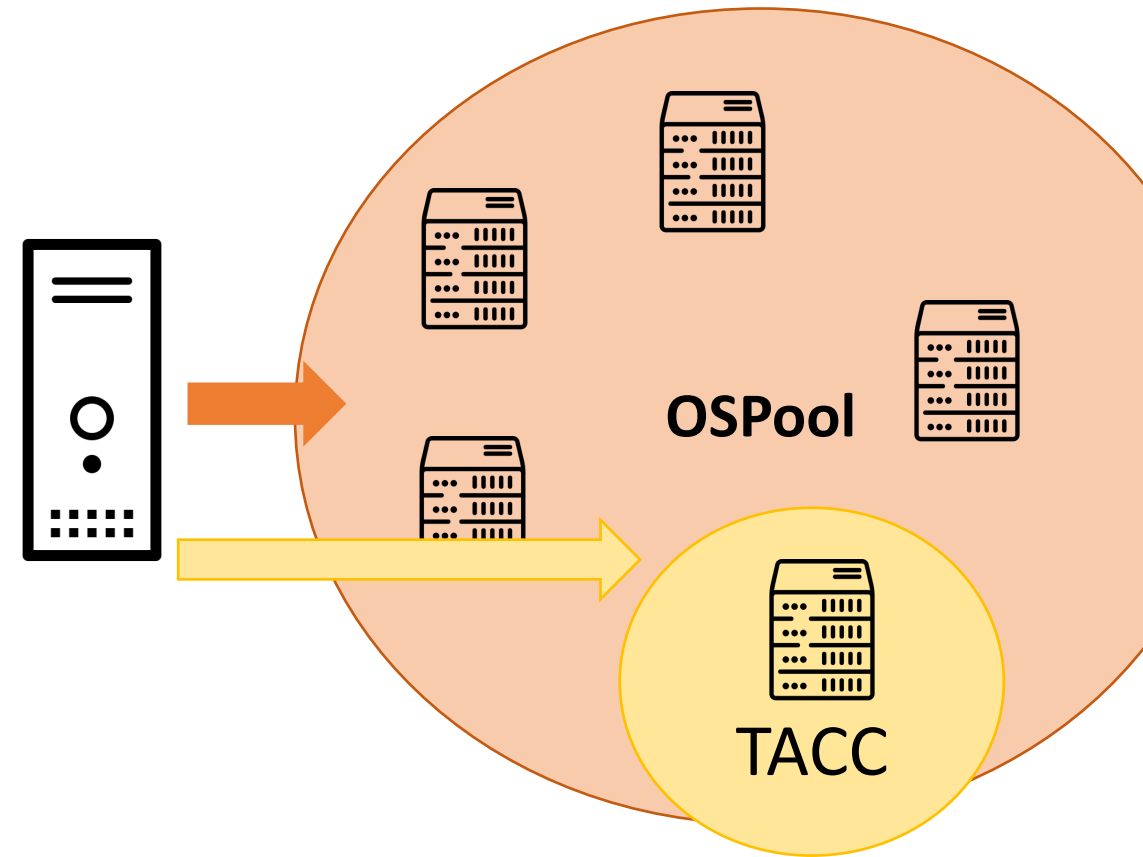
Dedicated Capacity: the PATh Facility

- What if your workloads are not a good fit for the capacity or restrictions of the OSPool, but you are still doing throughput-style work?
- Check out the PATh Facility – more details in the next talk.



Bring Your Own Capacity

- Do you want more specific capacity and have (for example) an ACCESS allocation?
- You can use a resource within your allocation from the Access Point, via a personal “annex”
- Talk to us if you are interested in this feature!



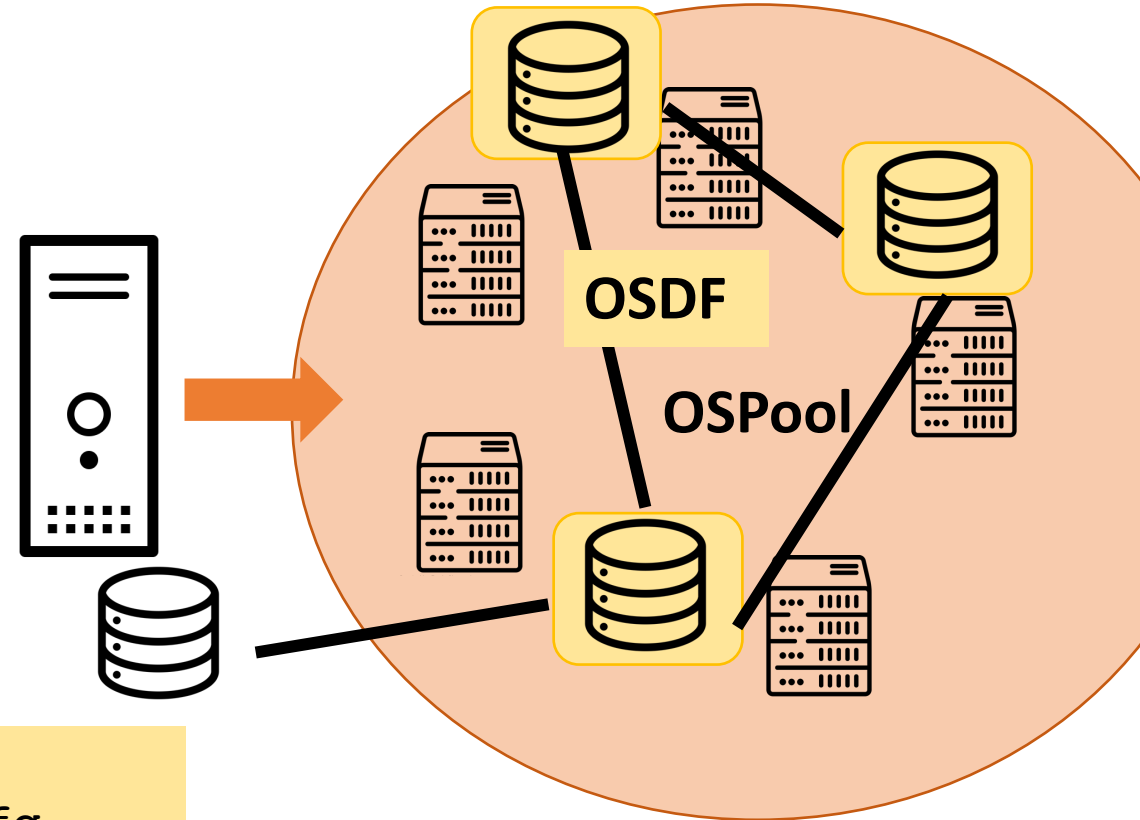
Staging Data

- As a distributed system, the OSPool does not rely on a shared file system.
- Instead, inputs must be fetched to the job's execution location, outputs moved back, usually included in the job description.
- Multiple mechanisms for data distribution
 - HTCSS default file transfer
 - Plugins to download data from web URLs or S3
 - Open Science Data Federation

Data Capacity: Open Science Data Federation

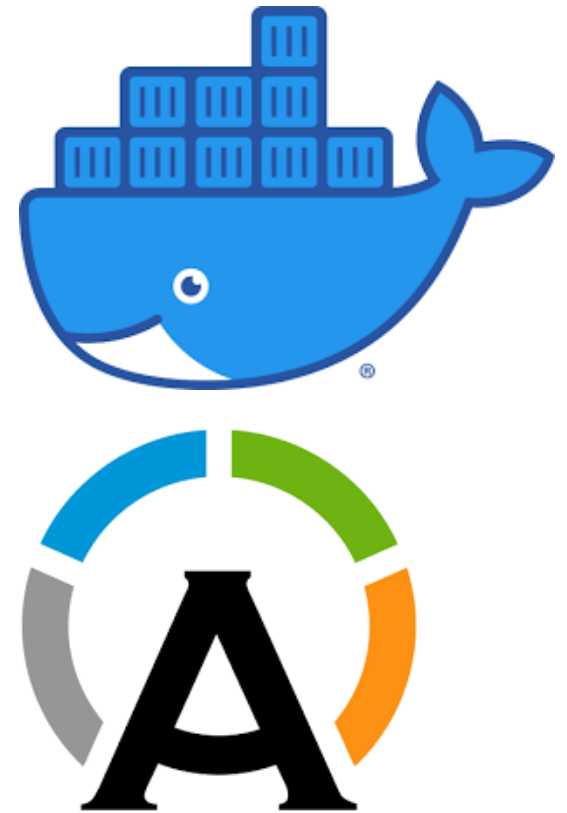
- The Open Science Data Federation is a network of data “origins” (usually associated with an Access Point) and caches for distributing job files
- Data can be fetched using a URL in the submit file or a command line tool.

```
transfer_input_files =  
    osdf:///ospool/data/alice/sample.fq
```



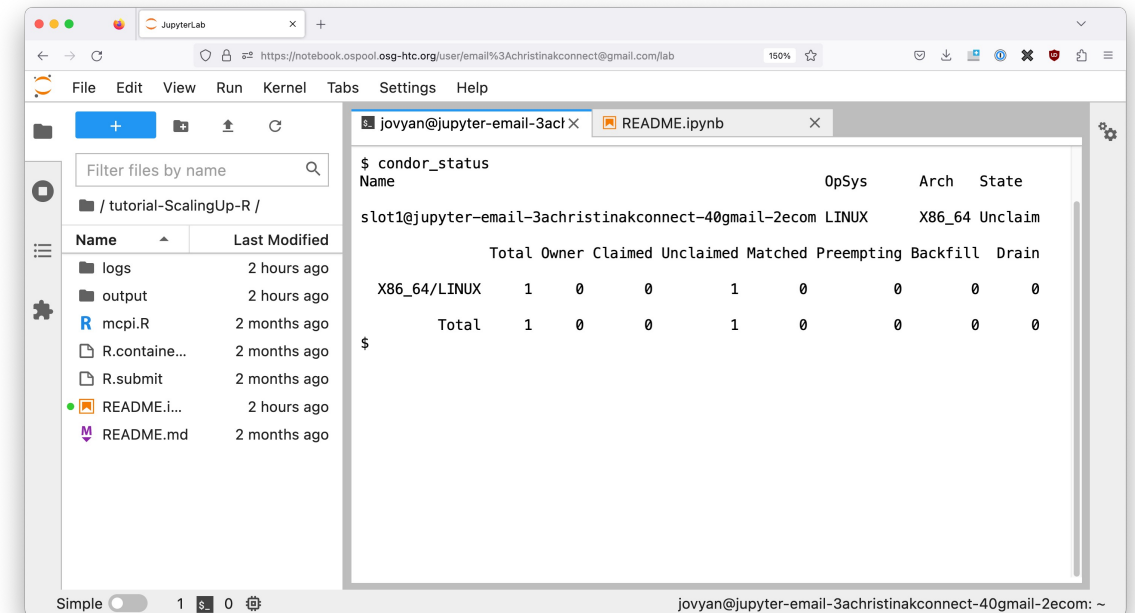
Software-Adjacent Services

- How to handle software environments? Need to create portable binaries/zip files or containers.
- HTCSS has container integration (both Docker and Singularity/Apptainer)
- Use previously described file transfer tools as delivery mechanisms
 - Also have a prototype container registry.



OSPool Notebooks

- Jupyter Hub interface to HTC resources
 - Access an HTC resource with just a browser
 - Includes terminal, file browser, notebook capabilities
 - Can be used with or without a full OSPool account
- Talk and demo coming up



<https://notebook.ospool.osg-htc.org>

Facilitation Services

- Individual orientation
- Email support
- Office hours
- Documentation
- Monthly training
- Feedback and feature requests to developers



Get Started

- Apply for an account and explore documentation

<https://portal.osg-htc.org/>

<https://portal.path-cc.io/>

- Come to a training (no account needed!)

https://portal.osg-htc.org/documentation/support_and_training/training/osgusertraining/

- Test drive a workflow using OSPool Notebooks

<https://notebook.ospool.osg-htc.org>

- Contact us!!

support@osg-htc.org

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 2030508. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.